

Hollow cone sieve for top

Vernon Barger, Peisi Huang

Department of Physics, University of Wisconsin, Madison, WI 53706, USA

A method of top tagging is introduced. Using the anti-kt algorithm to define jets, events with $n_j = 2$ fat jets of cone size $R = 1.5$ are decomposed into $R = 0.6$ sub-jets and retained if $n_j (R=0.6) \geq 4$. One pair of sub-jets reconstructs the W-mass and another jet is tagged as a b-jet, as necessary for hadronic or semileptonic events of $t\bar{t}$ origin. This 'hollow cone' method distinguishes the $t\bar{t}$ events from the light parton QCD dijet events.

The top quark is important in the Standard Model (SM) because of its large coupling to the Higgs field. Also, top is of great interest in new physics extensions of the SM because one or more new particles commonly decay into a single top quark or into top-quark pairs. Examples include SUSY [1], where the stop companion of top may decay to top and the gluino companion of the gluon may decay to top pairs, the Little Higgs Theory [2] where a heavy top may decay to top, the Randall-Sundrum model [3] where the decays of Kaluza-Klein gluons may lead to an enhancement of top signals. Thus, the study of top at the Tevatron and at the Large Hadron Collider (LHC) should be an excellent probe of new physics. We focus on the semileptonic channel (one $t \rightarrow b l \nu$ decay and one $t \rightarrow b q \bar{q}$ decay) and the hadronic channel (both tops decay to quarks); the leptonic channel with both tops decaying via $t \rightarrow b l \nu$ has a clean signal but a reduced rate. The expected $t\bar{t}$ event ratios in the SM are hadronic : semileptonic : leptonic = 0.322 : 1 : 3.13. Silicon vertex detectors have enabled the tagging of b-events with high efficiency ($\sim 50\%$) using the secondary vertex algorithm [4] at the Tevatron and b-tagging efficiencies of 60% [5], 65% [6] are estimated for the ATLAS and CMS detectors.

At the Tevatron collider, b-tagging and kinematical selection has been used to identify $t\bar{t}$ events in the fully hadronic channel [7]. For the semi-leptonic channel, the CDF collaboration employs b-tagging, jet multiplicity and an isolated electron or muon to do the event selection, and uses a matrix element method to tag a top [8]. The D0 collaboration selects events with exactly 4 jets and assigns the jets – 1 jet associated with the lepton and the other 3 jets with the hadronic top decay in a way that minimizes the reconstructed mass difference between the two top quarks [9].

The LHC is a top factory: in the SM about 8,000 top pairs should have been produced with the 47 pb^{-1} integrated luminosity already taken per detector at 7 TeV. Since the LHC center-of-mass energy is high compared to the top mass, the tops will typically be highly boosted, so that the decay products are close to each other, as illustrated in Figure 1. Thus, in the detector, at first sight, the top decay products may look like a fat jet instead of the several separate ones of which it is composed. Butterworth *et al.* look into subjets inside a fat jet [10] [11], to tag a Higgs or a W boson. One issue arises with the fat jet is that soft QCD contamination becomes relevant in the analysis of the jet substructure. Several techniques have been developed to deal with that complication. Jet filtering [10] keeps only several hard subjets, while jet trimming [12] tries to eliminate the soft ones. Jet pruning [13] avoids soft and large angle jet recombinations. Following these techniques, there are several algorithms to tag a top. Kaplan *et al.* [14] use the ratio of transverse momentum of the subjet to that of the original jet (the cut is 0.1 or 0.05) to find out which subjet is a “hard subjet” and which is from soft radiation, then apply a cut on invariant masses of the fat jet and 2 of the subjets, requiring them to be within the mass windows of top and W respectively. Also, a cut on helicity angle of the W, which is the angle between the reconstructed top and one of the W decay products in the reconstructed W rest frame [15], can be applied. Plehn *et al.* [16] use a mass drop criteria to find sub-jets (discussed below) and use a fixed mass window for top and W reconstruction, to tag tops with p_T above 200 GeV. There are also several top tagging algorithms using other kinematics variables, without tagging a W or a b, like two-body or multi-body kinematics [17], separation between lepton and hadronic activities of top [17] [18], and jet mass distribution [19].

Search strategy In the top reconstruction, to catch all the three main decay products of a top instead of a single fat jet, it is natural to use a large jet size, in other words, a large R parameter in the jet clustering, otherwise, several separate jets will be constructed instead of one. Here R is the jet size defined by $R = \sqrt{\eta^2 + \phi^2}$ where η is the pseudo-rapidity difference of two particles and ϕ is their azimuthal angle difference. The lower value of $R = 1.5$ for the top jet is suggested by the analyses of ref. [20]. Now consider a top pair. If a large $R \sim 1.5$ is used, it is likely that two jets will be constructed in the event (one from the top and the other from the \bar{t}), while more will be constructed using a small R . The light jets behave differently from the top jets – the number of reconstructed light jets does not vary with R . So after subtracting light jets from dijet events, the top contribution can be seen in the variation of the number of jets with cone size R . Figure 2 shows number of jets versus cone size R in events first selected as dijets with $R = 0.6$. The number of top pair events increases with R . Results are for the LHC at 7 TeV with 50 pb^{-1} ; the SM prediction for the number of top quark pairs is 8,000. Thus, we develop the “hollow cone” idea to tag top pairs. Consider the anti-kt algorithm as a “perfect cone” algorithm. When a larger cone size is used,

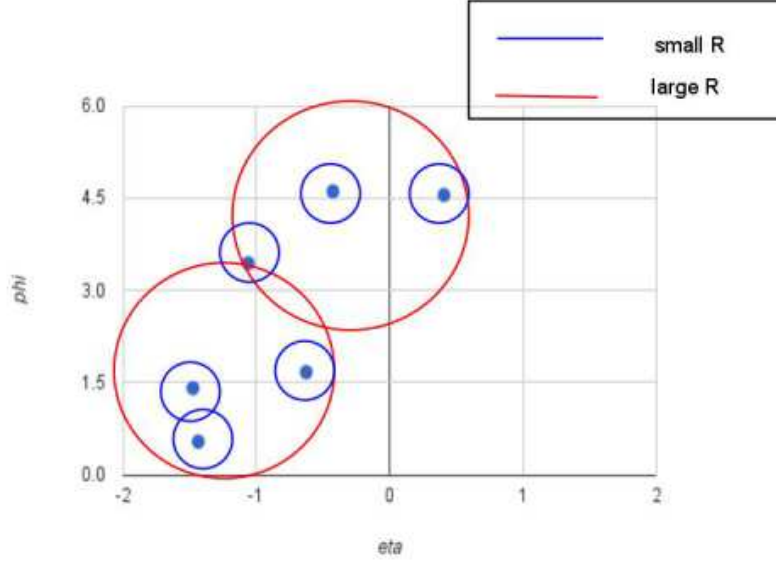
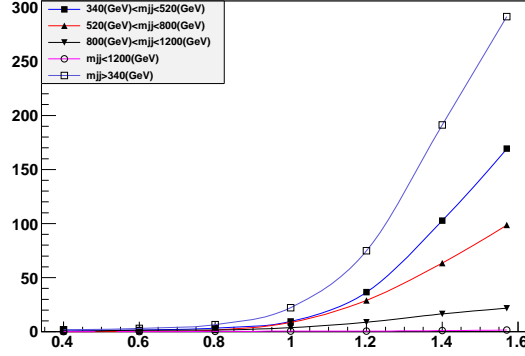
FIG. 1: subjects of a top in $\eta \phi$ plane

FIG. 2: number of top pairs in dijet events

both a $t\bar{t}$ event and a QCD dijet event will give two jets, when a smaller cone size is used, a $t\bar{t}$ event will have more jets while a QCD dijet event still have two. This means, for a fat jet with a large cone size, after subtracting the small cone in the middle, if there something remains in the hollow cone, the jet is likely to be a top jet, if there is nothing in the hollow cone, it is likely to be a light quark jet or a gluon jet.

Our top tagging algorithm proceeds in the following steps, trying to separate top pairs from the QCD dijet events:

- 1 Reconstruct jets using the anti-kt [21] jet algorithm with $R = 1.5$ to obtain a set of jets. The number of jets is n_{jets} .
- 2 Redo the jet reconstruction, with $R = 0.6$ (or $R = 0.7$), following recent works of ATLAS [22] and CMS [23], to obtain another set of jets.
- 3 Keep the event as a $t\bar{t}$ candidate if $n_{jets, R=1.5} = 2$ and $n_{R=0.6} > 2$.
- 4 Go into the 2 jets reconstructed in step 1, find all the subjects for each fat jet, using the method described in [16], as follows. For a fat jet of invariant mass of m_j , undo the last step of jet clustering to obtain two jets j_1 and j_2 , with

invariant masses m_{j1} and m_{j2} ($m_{j1} > m_{j2}$). If $m_{j1} < 0.9 m_j$, keep both j_1 and j_2 , otherwise, keep only j_1 to add to the subjet list and decompose further. Add j_i to the jet substructure list if $m_{ji} < 30$ GeV, otherwise decompose j_i iteratively. If the total number of subjets is less than 4, reject the event, because a hadronic top and one semileptonic top should give 4 subjets in total, and two hadronic tops will result in 6 subjets.

5 See whether there is a W inside either of the 2 fat jets, if not, reject the event. To do this, look into a fat jet and iterate over all of the 2 subjets configurations. After the jet filtering [24], if the invariant mass of the 2 subjets falls in the window of 65 GeV to 95 GeV, tag that configuration as a W. A similar method of W-tagging is discussed in [14], [16], [25], and [26].

6 See whether there is a b inside either of the 2 jets by b-tagging; if not, reject the event.

Any event that survives the above sequence is tagged as a $t\bar{t}$ event.

Signal and Background The main backgrounds here are $Wb\bar{b}$ and $Zb\bar{b}$. Since there will be b jets in both cases, and the Z mass is close to W mass, these two backgrounds are indistinguishable in their hadronic decay channels. Others background are the QCD dijets from light quarks and gluons and QCD multi-jet events. QCD dijets events will be gotten rid of by the “hollow cone” cut. QCD trijet events will be eliminated by requiring 4 or more subjets.

We generate the parton level events with MadEvent[27], then use Pythia[28] to do the parton shower and hadronization. FastJet[29] is used to reconstruct jets and analyze the jet substructure. Then we apply the following cuts in sequence :

cut 1 : The “hollow cone” cut. Require $n_{jets} = 2$ and $n_{veto} > 2$.

cut 2 : Total number of subjets ≥ 4 .

cut 3 : A hadronic W can be tagged.

cut 4 : A b jet can be tagged.

The cut flow table of signal and background is in Table I. We assume a 0.5 b-tagging efficiency here. The recon-

| | cross section(pb) | cut 1(pb) | cut 2(pb) | cut 3(pb) | cut 4(pb) |
|-------------|-------------------|-----------|-----------|-----------|-----------|
| $t\bar{t}$ | 100.00 | 12.63 | 7.59 | 5.39 | 4.05 |
| $Wb\bar{b}$ | 239.52 | 63.93 | 1.40 | 0.20 | 0.18 |
| $Zb\bar{b}$ | 124.81 | 23.55 | 1.20 | 0.57 | 0.43 |

TABLE I: cut flow table for signal and background

structed W mass and top mass can be seen in figure 3 and 4.

The resulting ratio of hadronic tops to semileptonic tops is 2.81, which is consistent with the ratio of decay branching

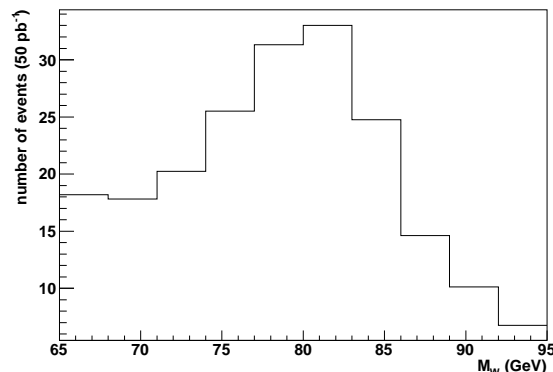


FIG. 3: reconstructed W mass

fractions of 3.13. The transverse momentum distribution of the tagged t and \bar{t} is shown in figure 5 and compared with the transverse momentum of the QCD dijet. It shows that the method is picking top jets instead of light jets, and also demonstrates that top jets with relatively low p_t can be tagged. Former top tagging techniques require the top to be harder than 200 GeV [16].

Figure 6 and 7 show the distance in R of W with the nearest b jet for $t\bar{t}$ and $Wb\bar{b}$. The $t\bar{t}$ events are more likely to have a b jet inside the W, while in $Wb\bar{b}$ the W and b are farther apart, as expected.

Outlook This method can be also used in identifying new physics that has a top in the final state. The tagged top can serve as a distinguishable signature. Also, the “hollow cone” can be used for discovering new, relatively heavy and boosted particles at the LHC after taking SM top contribution into consideration. If the new particle has a mass

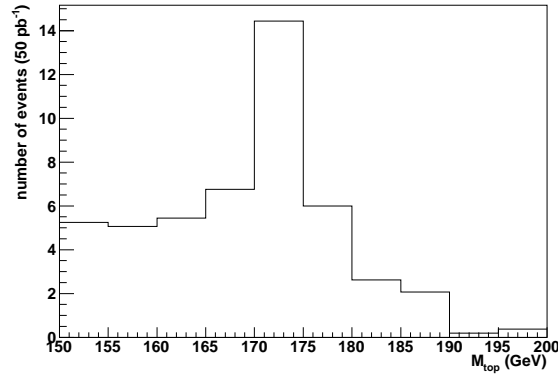
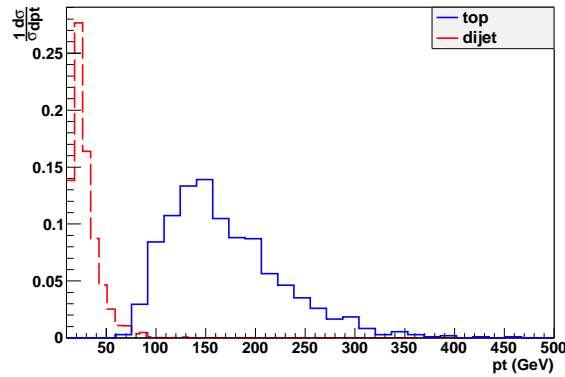


FIG. 4: reconstructed top mass

FIG. 5: Normalized p_T distributions for top and dijets

is not very close to the top mass, the mass of the new particle can be reconstructed by the invariant mass of the filtered subjects.

-
- [1] Savas Dimopoulos and Howard Georgi. Softly Broken Supersymmetry and SU(5). *Nucl. Phys.*, B193:150, 1981.
 - [2] Nima Arkani-Hamed, Andrew G. Cohen, and Howard Georgi. Electroweak symmetry breaking from dimensional deconstruction. *Phys. Lett.*, B513:232–240, 2001.
 - [3] Lisa Randall and Raman Sundrum. A large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999.
 - [4] V. M. Abazov et al. Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ -TeV. *Phys. Rev. Lett.*, 100:192004, 2008.
 - [5] G. Aad et al. Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics. 2009.
 - [6] Roberta Volpe and for the CMS collaboration. Probing the Heavy Flavor Content in $t\bar{t}$ Events and Using $t\bar{t}$ Events as a Calibration Tool at CMS. 2009.
 - [7] Daniel Wicke. Top pair production cross-section measurement in the all- hadronic channel at CDF and D0. *Int. J. Mod. Phys.*, A20:3183, 2005.
 - [8] T. Aaltonen et al. Top Quark Mass Measurement in the Lepton + Jets Channel Using a Matrix Element Method and in situ Jet Energy Calibration. *Phys. Rev. Lett.*, 105:252001, 2010.
 - [9] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.
 - [10] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. Jet substructure as a new Higgs search channel at the LHC. *Phys. Rev. Lett.*, 100:242001, 2008.
 - [11] J. M. Butterworth, B. E. Cox, and Jeffrey R. Forshaw. WW scattering at the CERN LHC. *Phys. Rev.*, D65:096014, 2002.
 - [12] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jet Trimming. *JHEP*, 02:084, 2010.
 - [13] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Techniques for improved heavy particle searches with

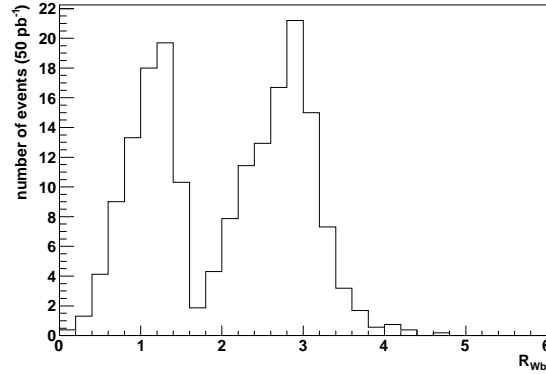


FIG. 6: Distance in R between the W -boson and the nearest b jet in $t\bar{t}$ events.

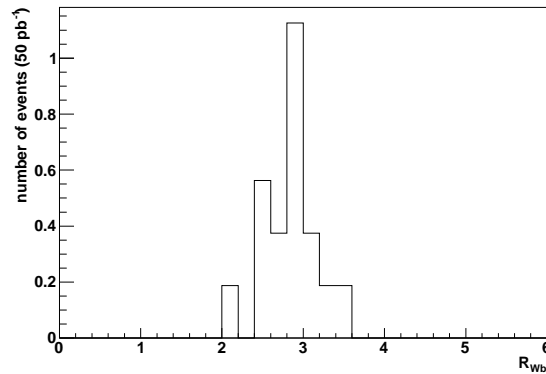


FIG. 7: Distance in R between W and the nearest b jet in $Wb\bar{b}$ events.

jet substructure. *Phys. Rev.*, D80:051501, 2009.

- [14] David E. Kaplan, Keith Rehermann, Matthew D. Schwartz, and Brock Tweedie. Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks. *Phys. Rev. Lett.*, 101:142001, 2008.
- [15] Thorsten Chwalek. Measurement of the W^- boson helicity fractions in top- quark decays at CDF. 2007.
- [16] Tilman Plehn, Gavin P. Salam, and Michael Spannowsky. Fat Jets for a Light Higgs. *Phys. Rev. Lett.*, 104:111801, 2010.
- [17] Jesse Thaler and Lian-Tao Wang. Strategies to Identify Boosted Tops. *JHEP*, 07:092, 2008.
- [18] Keith Rehermann and Brock Tweedie. Efficient Identification of Boosted Semileptonic Top Quarks at the LHC. 2010.
- [19] Leandro G. Almeida, Seung J. Lee, Gilad Perez, Ilmo Sung, and Joseph Virzi. Top Jets at the LHC. *Phys. Rev.*, D79:074012, 2009.
- [20] Tilman Plehn, Michael Spannowsky, Michihisa Takeuchi, and Dirk Zerwas. Stop Reconstruction with Tagged Tops. *JHEP*, 10:078, 2010.
- [21] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *JHEP*, 04:063, 2008.
- [22] G. Aad et al. Search for Quark Contact Interactions in Dijet Angular Distributions in pp Collisions at $\sqrt{s} = 7$ TeV Measured with the ATLAS Detector. *Phys. Lett.*, B694:327–345, 2011.
- [23] Vardan Khachatryan et al. Search for Dijet Resonances in 7 TeV pp Collisions at CMS. *Phys. Rev. Lett.*, 105:211801, 2010.
- [24] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. Jet substructure as a new Higgs search channel at the Large Hadron Collider. 2008.
- [25] Yanou Cui, Zhenyu Han, and Matthew D. Schwartz. W-jet Tagging: Optimizing the Identification of Boosted Hadronically-Decaying W Bosons. 2010.
- [26] Vardan Khachatryan et al. A cambridge-aachen (c-a) based jet algorithm for boosted top-jet tagging. Jul 2009.
- [27] Johan Alwall et al. MadGraph/MadEvent v4: The New Web Generation. *JHEP*, 09:028, 2007.
- [28] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.
- [29] Matteo Cacciari and Gavin P. Salam. Dispelling the N^3 myth for the k_t jet-finder. *Phys. Lett.*, B641:57–61, 2006.